



OPEN ACCESS

EDITED BY

Takuya Takahashi,
Yokohama City University, Japan

REVIEWED BY

Naoki Nakagawa,
Kyoto University, Japan
Robin Panneton,
Virginia Tech, United States
Iris-Corinna Schwarz,
Stockholm University, Sweden

*CORRESPONDENCE

Sho Tsuji

✉ sho.tsuji@ens.psl.eu

Irena Lovčević

✉ irena.lovcevic@ircn.jp

RECEIVED 31 December 2024

ACCEPTED 12 May 2025

PUBLISHED 30 May 2025

CITATION

Tsuji S, Alonso F, Hagihara H,
Kimura N, Polka L and Lovčević I (2025) Facial
affect and arousal as a complement to gaze
measures in infant speech sound perception
studies.

Front. Neural Circuits 19:1553918.

doi: 10.3389/fncir.2025.1553918

COPYRIGHT

© 2025 Tsuji, Alonso, Hagihara, Kimura, Polka
and Lovčević. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Facial affect and arousal as a complement to gaze measures in infant speech sound perception studies

Sho Tsuji^{1,2*}, Fernanda Alonso³, Hiromichi Hagihara^{2,4},
Nanako Kimura⁵, Linda Polka³ and Irena Lovčević^{2*}

¹Laboratoire de Sciences Cognitives et Psycholinguistique, ENS, EHESS, CNRS, PSL University, Paris, France, ²International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, Tokyo, Japan, ³School of Communication Sciences and Disorders, Centre for Research on Brain, Language and Music, McGill University, Montreal, QC, Canada, ⁴Graduate School of Human Sciences, The University of Osaka, Suita, Japan, ⁵Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

This study explores infant facial expressions during visual habituation to investigate perceptual attunement to native and non-native speech sounds. Using automated facial affect analysis based on Facial Action Units, we analyzed valence, arousal, positive affect, and negative affect during the experiment. Valence and arousal decreased with habituation, while positive affect increased, with differences between native and non-native stimuli. Facial affect showed links to discrimination outcomes, with better native discrimination linked to reduced negative affect. These findings highlight the potential of facial expression analysis as a complementary tool to gaze-based measures in early language development research.

KEYWORDS

speech perception, infancy, perceptual attunement, facial affect, automatic facial analysis

1 Introduction

Human infants' ability to learn the phonemes of their native language relies on neural mechanisms of perceptual attunement. Initially capable of discriminating a wide range of phoneme contrasts across world languages, infants typically show a decline in sensitivity to non-native phonemes and an improvement in native phoneme discrimination within the first year of life (Werker and Tees, 1984). This experience-dependent tuning reflects synaptic pruning and neural plasticity during critical developmental periods and is positively linked to later language skills (Werker and Hensch, 2015).

To study perceptual attunement, infants are typically habituated to a specific speech sound, and dishabituation to a novel sound is measured via looking times. This method robustly demonstrates perceptual attunement across languages and cultures (Tsuji and Cristia, 2014).

However, perceptual attunement is not a uniform process, and factors such as the richness of the linguistic and social environment shape its trajectory. For instance, bilingual infants may take longer to attune their speech sound categories compared to monolingual infants (Bosch and Sebastián-Gallés, 2003), and socially enriched learning environments enhance infants' acquisition of speech sound contrasts (Kuhl et al., 2003). These findings suggest that attunement is influenced by the interplay of linguistic exposure, social interaction, and cognitive development.

Given these multifaceted influences, it could be insightful for discrimination measures to also capture the richness of infants' responses. While the conventional looking time measure is invaluable, it may miss other signals, such as facial expressions, that provide additional

insight into the affective and cognitive processes underlying attunement. Video recordings of infants' faces in habituation studies offer untapped potential for such analyses.

Advances in automated analysis tools, such as FaceReader and its infant-specific adaptation BabyFaceReader (Noldus Information Technology, 2021), enable the systematic extraction of facial expressions using the Facial Action Coding System (Ekman and Friesen, 1978). This framework quantifies affective states by identifying discrete facial muscle movements, or action units (AUs). Affective expressions can be classified within a bidimensional framework of valence and arousal: valence reflects the intrinsic positivity or negativity of a state, while arousal represents its intensity or activation level. During habituation, arousal is expected to decrease and then increase in response to a novel stimulus (e.g., Thompson, 2009), while valence may vary depending on factors such as age and stimulus complexity (e.g., Hunter and Ames, 1988). In addition to valence and arousal, specific AU configurations allow the assessment of negative affect (unpleasant emotions) and positive affect (pleasant emotional states), which have been found to be closer aligned to manual coding of affect (Zaharieva et al., 2024) and which can provide a more granular view of infants' affective state.

By analyzing these dimensions, we aim to enrich our understanding of the affective and attentional responses associated with perceptual attunement. Our study focuses on Japanese infants' developing discrimination of two phonemic contrasts: a native vowel length contrast and a non-native r-l contrast. Both have been documented to exhibit characteristic patterns of perceptual attunement, with studies using the same native contrast and testing method showing improved discrimination between 6 and 12 months. Six-month-olds failed to discriminate the contrast, while 12-month-olds succeeded (Lovčević and Tsuji, 2024; Sato et al., 2010). Results for 9-month-olds were mixed, with successful discrimination reported by Lovčević and Tsuji (2024) but not Sato et al. (2010). For the non-native r-l contrast, studies have consistently shown a decline in discrimination over the first year. Japanese infants aged 6–8 months could discriminate the contrast, but this ability diminished by 10–12 months (Kuhl et al., 2006; Tsushima et al., 1994). Using the same contrast and method as in the present study, Lovčević and Tsuji (2024) reported successful discrimination at 9 months but not at 12 months.

The current dataset includes longitudinal video data on infants' discrimination of the native and non-native contrast at 9 and 12 months. By analyzing facial expressions during speech sound discrimination, we aim to reveal the potential of integrating this measure into the study of perceptual attunement.

2 Methods

2.1 Participants

All experiments were conducted within-subjects, but not all combinations of infant and data points could be retained. For looking time data in the native speech sound discrimination experiment, data from 31 9-month-old infants (range 246–297 days, 16 female) and 33 12-month-old infants (range 367–395 days, 18 female) were included. A further five 9-month-old infants and one 12-month-old infant were excluded from analysis due to not completing the task. For the non-native speech sound discrimination experiment, data

from 29 9-month-old infants (range 246–297 days, 14 female), and 29 12-month-old infants (range 365–395 days, 15 female) were included. Six 9-month-old and two 12-month-old infants were excluded from analysis due to not completing the task or not habituating.

For the facial expression analyses, we excluded participants who were missing video recordings due to experimenter or equipment error or missing parental approval. The sample for native speech perception included 20 9-month-old infants (range 246–293, 14 female) and 24 12-month-old infants (range 367–395, 12 female). As for non-native speech perception, the final sample included 18 9-month-old infants (range 246–293, 12 female), and 22 12-month-old infants (range 365–395, 11 female).

All infants were monolingual Japanese recruited in Tokyo, and caregivers gave written informed consent for participation and data reuse before the experiment. The protocol was approved by the local IRB committee (name anonymized for submission).

2.2 Stimuli

Stimuli consisted of native and non-native speech sounds, and were identical to those in Lovčević and Tsuji (2024). The native contrast consisted of non-words /mana/ and /ma:na/, differing in vowel duration. The /ma:na/ stimuli were recorded by a female Japanese speaker, while /mana/ stimuli were generated by shortening the steady part of /a:/ to control acoustic differences. The non-native contrast used the English words “right” (/raɪt/) and “light” (/laɪt/), unfamiliar to Japanese infants, and recorded by a female Canadian English speaker. Stimuli were recorded in infant-directed register and presented in lists of 10 repetitions with ~1.5-s intervals. Since this study involved affect-related facial expressions, six adults rated the valence of the stimuli (following Kitamura and Burnham, 2003, but without low-pass filtering). While adult ratings may not fully capture infant perception – particularly as non-native stimuli consist of real words, potentially influencing judgments – they provide some insight into valence differences. Median valence was higher for native (1, IQR = 3) than non-native stimuli (0, IQR = 1.25), with ordinal regression confirming this difference as significant (Odds ratio = 0.452, $p = 0.035$). Since native stimuli were disyllabic and thereby longer than non-native monosyllables, they might have provided more possibility to express affect.

2.3 Procedure

Each experimental session consisted of one experiment using native, and one experiment using non-native stimuli. Each infant participated in two sessions (at 9 and 12 months). Experiments took place in a dimly lit, soundproof lab and were controlled using Habit X 2.0 software (Oakes et al., 2019) on a Windows computer. Visual stimuli were displayed on a 23-inch screen, and audio stimuli were delivered via forward-facing speakers below the screen. Infants sat on their caregiver's lap, while caregivers wore noise-canceling headphones playing music to mask the auditory stimuli. An experimenter observed the infant through a video camera and recorded their gaze to the screen by pressing a key on the computer keyboard, with key presses logged for analysis. Infants' gaze was also recorded on video.

The study used a modified visual habituation paradigm (Stager and Werker, 1997) with four phases: pre-test, habituation, test, and post-test. Before each trial, an attention getter was presented, and as soon as the infant looked to the screen, the experimenter pressed the key on the experimental computer keyboard to start the next trial. Pre- and post-test consisted of one trial each, featuring a rotating wheel animation paired with the audio stimulus “panta.” Comparing looking behavior in these phases assessed potential fatigue or disinterest.

Presentation order of the native and non-native experiment was counterbalanced across infants, and infants took a short break between experiments. Within the native or non-native experiment, half of the infants were habituated to one type of speech sound (/mana/ or /ma:na/ for the native, and /aɪt/ or /laɪt/ for the non-native experiment).

In each experiment, during habituation, stimulus lists representing one type of speech sound were presented accompanied by a static red-and-black checkerboard visual. Habituation ended when infants met the criterion (average looking time in a four-trial sliding block <60% of the first block) or completed 28 trials.

In the test phase, infants completed four trials: two “switch” trials with the respective other speech sound, followed by two “same” trials matching the habituation stimulus. For example, infants habituated to /ma:na/ heard /mana/ in switch trials and /ma:na/ in same trials.

2.4 Data preprocessing

To analyze infant gaze behavior, we extracted logged keypress data, indicating looks to and away from the screen during each trial.

Facial expressions were analyzed using BabyFaceReader (Noldus Information Technology, 2021), which detects infant faces, maps facial landmarks with a mesh, and analyzes gaze, head orientation, and facial movements frame-by-frame. BabyFaceReader then generates outputs for all AUs and combines these to code specific expressions, in our case valence and arousal. Valence, ranging from -1 (maximum negative emotion) to $+1$ (maximum positive emotion), reflects overall emotional tone, while arousal, ranging from 0 (minimum) to 1 (maximum), indicates intensity. In addition to these valence scores (which combine many AUs), we analyzed positive and negative affect using selected AUs. Negative affect, derived from AU3 and AU4 (brow lowering), and positive affect, represented by AU12 (lip corner puller), range from 0 (low) to 1 (high) (Maroulis, 2018; Noldus Information Technology, 2021), and were added to the analyses for two reasons. First, a recent study by Zaharieva et al. (2024) found a closer alignment between manual and automated coding of positive affect when using AU12 compared to using the composite valence score. Similarly, they also found good performance when using AU3 + AU4 to categorize negative affect. Second, there are open-source tools that measure these specific action units, whereas BabyFaceReader is costly and hence, not widely available.

The videos were recorded continuously during each experiment and were not synced with the experimental software, and therefore did not include trial start and end times. To address this, human annotators manually coded the onset of experimental phases (pretest, habituation, test, and post-test) in the videos based on sound onsets. This timing information was then integrated with trial onset data from the experimental software and aligned with the BabyFaceReader output.

2.5 Analysis plan

We first analyzed gaze behavior to assess perceptual attunement, following Lovčević and Tsuji (2024). Linear mixed-effects models were fitted separately for the native and non-native experiments in each age group using the *lmer* function from the lme4 package (Bates et al., 2015) in R (R Core Team, 2013). The models included Trial (same, switch), Habituation Stimulus (native: /mana/, /ma:na/, non-native: /aɪt/, /laɪt/), and their interaction as fixed effects, with random intercepts for participants and infants’ looking time as the dependent variable. Significance was tested using ANOVAs with Satterthwaite’s method via the *anova* function of the *lmerTest* package (Kuznetsova et al., 2017).

We next conducted three sets of confirmatory analyses on facial expressions, each comprising four linear mixed effects models with valence, arousal, negative affect, or positive affect as the dependent variable. Categorical predictors common to all analyses were Age (9, 12 months) and Nativeness (native, non-native), and were sum-coded. Models included random intercepts for participants. Significance was tested with Type III ANOVAs (*car* package, Fox and Weisberg, 2019), and *post hoc* pairwise Bonferroni-corrected comparisons were performed (*emmeans*, Lenth, 2019). Visualizations used *ggplot2* (Wickham et al., 2016).

First, to track changes in facial expressions during the habituation phase, we assessed the effect of Age, Nativeness, Trial Number, and their interactions. The continuous predictor Trial Number (1–8) reflected the habituation trial number, where, to avoid sparse data, we excluded trials beyond the point where trial counts dropped below 50% of the initial count (trial 8). This predictor was centered.

Second, we examined differences in facial expressions between same and switch trials in the test phase, focusing on subsets likely to show behavioral discrimination. A group-based subset included 12-month-olds exposed to native and 9-month-olds exposed to non-native contrasts, as these groups are expected to show discrimination (Lovčević and Tsuji, 2024). An individual-based subset included all infants with an above-chance Novelty Score, a score reflecting the degree to which infants look more to trials with the novel compared to the known (habituated to) stimulus. The Novelty Score is calculated as looking time to switch trials divided by total looking time during switch and same trials [as in Arterberry and Bornstein (2002)]. Models included Age, Nativeness, and Phase (same, switch).

Third, to assess whether facial expressions showed any relationship to infants’ discrimination performance, we conducted an exploratory analysis predicting facial expressions during habituation based on infants’ Novelty Score. Facial expression values during habituation were modeled as a function of Novelty Score and Age.

3 Results

3.1 Speech perception

3.1.1 Native speech perception

Infants both at 9 and 12 months demonstrated significant discrimination [9 m: $F(1, 87) = 5.03, p < 0.05$; switch: $M = 6,655$ ms, $SD = 2,952$, same: $M = 5,288$ ms, $SD = 2,993$; 12 m: $F(1, 82) = 5.17, p < 0.05$; switch: $M = 5,926$ ms, $SD = 2,897$, same: $M = 4,709$ ms,

SD = 2052]. We note that another group of 6-month-olds (not included in this manuscript) did not show evidence of successful discrimination; thus, these findings align with patterns found in the perceptual attunement literature.

3.1.2 Non-native speech perception

Infants at both 9 and 12 months showed no significant discrimination [9 m: $F(1, 84) = 3.45, p = 0.07$; switch: $M = 4,946$ ms, $SD = 3,460$, same: $M = 4,078$ ms, $SD = 1966$; 12 m: $F(1, 85) = 0.40, p = 0.53$; switch: $M = 3,892$ ms, $SD = 2071$, same: $M = 4,197$ ms, $SD = 2,474$]. Based on previous work using the same stimuli (Lovčević and Tsuji, 2024) we would have expected discrimination at 9 months; however, the p -value suggests a tendency in a direction consistent with perceptual attunement.

3.2 Facial expressions

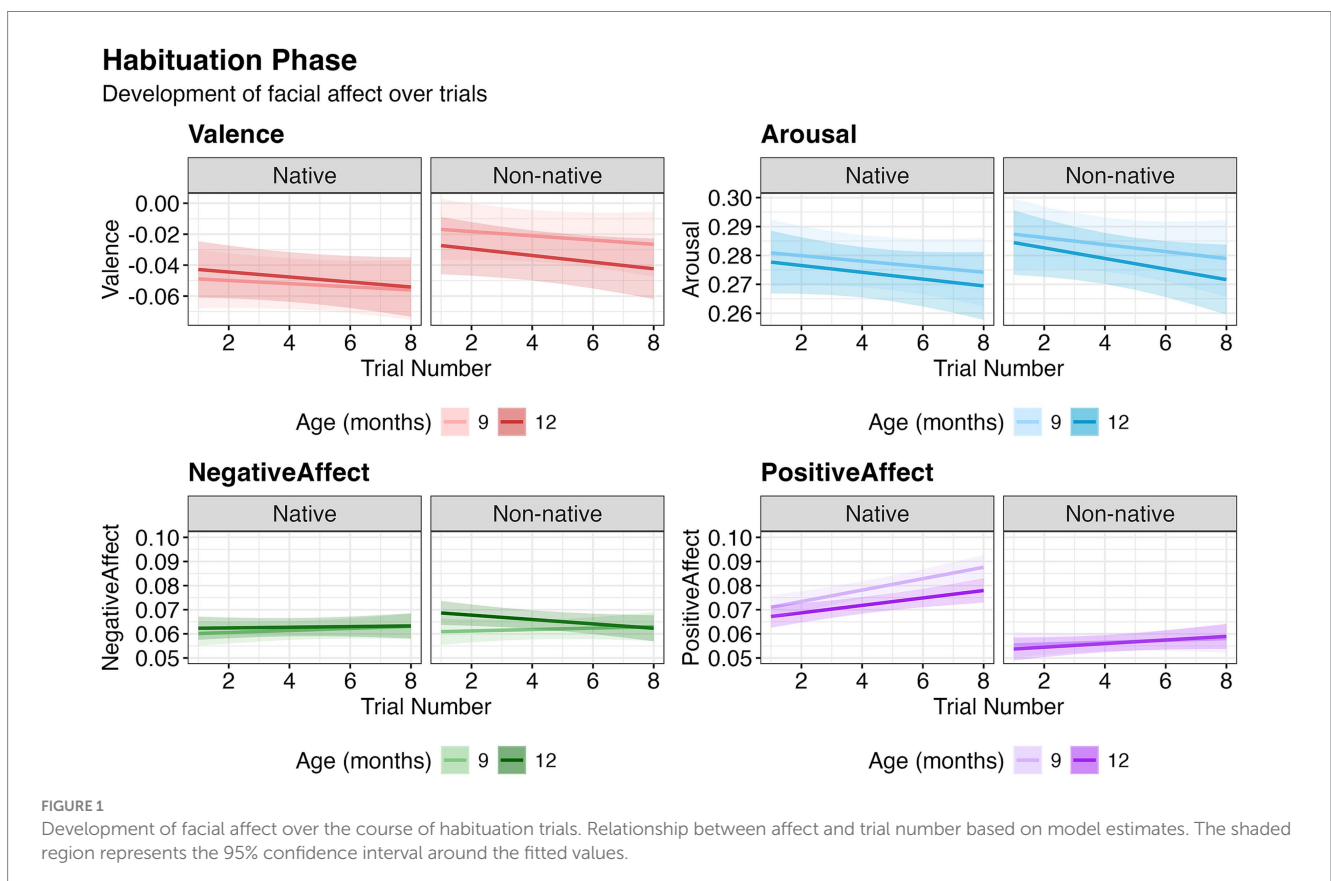
3.2.1 Habituation phase

Valence decreased significantly with increasing Trial Number [$\chi^2(1) = 4.19, p = 0.043$]. Nativeness had a significant effect [$\chi^2(1) = 43.31, p < 0.001$], with valence higher in response to non-native ($M = -0.030, SD = 0.039$) compared to native ($M = -0.050, SD = 0.069$) stimuli. Age interacted with Nativeness [$\chi^2(1) = 6.31, p = 0.012$], and post-hoc paired comparisons showed that the Age difference was significant for non-native [$t(590) = 2.36, p = 0.019$], but not for native stimuli [$t(590) = -0.82, p = 0.414$]. Arousal attenuated with increasing Trial Number [$\chi^2(1) = 5.28,$

$p = 0.022$], and was higher for non-native ($M = 0.280, SD = 0.030$) than native ($M = 0.276, SD = 0.040$) stimuli [$\chi^2(1) = 4.16, p = 0.041$]. Negative affect did not change over trials [$\chi^2(1) = 0.01, p = 0.091$]. Positive affect increased with increasing Trial Number [$\chi^2(1) = 22.92, p < 0.001$]. Age had a significant effect [$\chi^2(1) = 6.65, p = 0.010$], with higher positive affect at 9 months ($M = 0.069, SD = 0.021$) than at 12 months ($M = 0.066, SD = 0.016$). Positive affect was higher for native ($M = 0.076, SD = 0.022$) than for non-native ($M = 0.057, SD = 0.006$) stimuli [$\chi^2(1) = 262.51, p < 0.001$]. Trial number interacted with nativeness ($\chi^2(1) = 6.97, p = 0.008$), but post-hoc analyses showed that positive affect increased significantly with increasing trial number for both native [$\chi^2(1) = 22.51, p < 0.001$] and non-native [$\chi^2(1) = 17.29, p < 0.001$] stimuli separately. Age interacted with nativeness [$\chi^2(1) = 6.62, p = 0.010$], and follow-up paired comparisons revealed that the effect of age was only significant for native [$t(595) = 3.76, p < 0.001$], but not non-native stimuli [$t(595) = 0.25, p = 0.803$] (Figure 1). No other effects were significant. These findings suggest that facial affect reflects the habituation process, including decreases in valence and arousal and increases in positive affect, with variations across native/non-native stimuli and infant age.

3.2.2 Test phase

No significant differences were found between same and switch trials across facial expressions in the group-based analysis [valence: $\chi^2(1) = 0.85, p = 0.356$; arousal: $\chi^2(1) = 0.91, p = 0.338$; negative affect: $\chi^2(1) = 0.08, p = 0.784$; positive affect: $\chi^2(1) = 0.05, p = 0.822$]. Valence was significantly influenced by group ($\chi^2(1) = 7.38,$



$p = 0.007$), with 9-month-olds exposed to non-native stimuli ($M = -0.031$, $SD = 0.046$) showing higher valence than 12-month-olds exposed to native stimuli ($M = -0.047$, $SD = 0.061$). Positive affect also was significantly influenced by group [$\chi^2(1) = 11.75$, $p < 0.001$], with 12-month-olds exposed to native stimuli ($M = 0.076$, $SD = 0.020$) showing higher positive affect than 9-month-olds exposed to non-native stimuli ($M = 0.068$, $SD = 0.031$).

Similarly, no significant differences between same and switch trials were found in the individual-based analysis [valence: $\chi^2(1) = 0.05$, $p = 0.823$; arousal: $\chi^2(1) = 0.98$, $p = 0.321$; negative affect: $\chi^2(1) = 0.140$, $p = 0.708$; positive affect: $\chi^2(1) = 1.72$, $p = 0.189$]. Positive affect was significantly influenced by Nativeness [$\chi^2(1) = 20.08$, $p < 0.001$], with infants exposed to native stimuli ($M = 0.078$, $SD = 0.018$) showing higher positive affect infants exposed to non-native stimuli ($M = 0.059$, $SD = 0.048$). No other effects were significant (Figure 2). These results indicate that facial affect does not differ as a function of differences in behavioral responses indicating discrimination.

3.2.3 Relationship with discrimination scores

When predicting facial expressions during habituation using gaze-based discrimination scores from the test phase, no significant relationships for valence [$\chi^2(1) = 0.31$, $p = 0.580$], arousal [$\chi^2(1) = 0.26$, $p = 0.608$] or positive affect [$\chi^2(1) = 0.86$, $p = 0.354$] were observed. There was an interaction between the novelty score and nativeness for negative affect [$\chi^2(1) = 6.15$, $p = 0.0130$], and post-hoc analyses revealed that negative affect decreased with higher novelty scores for native speech sounds [$\chi^2(1) = 9.89$, $p = 0.002$], but not for non-native speech sounds [$\chi^2(1) = 0.19$, $p = 0.665$] (Figure 3). These preliminary findings suggest a potential link between facial expressions and language measurement performance.

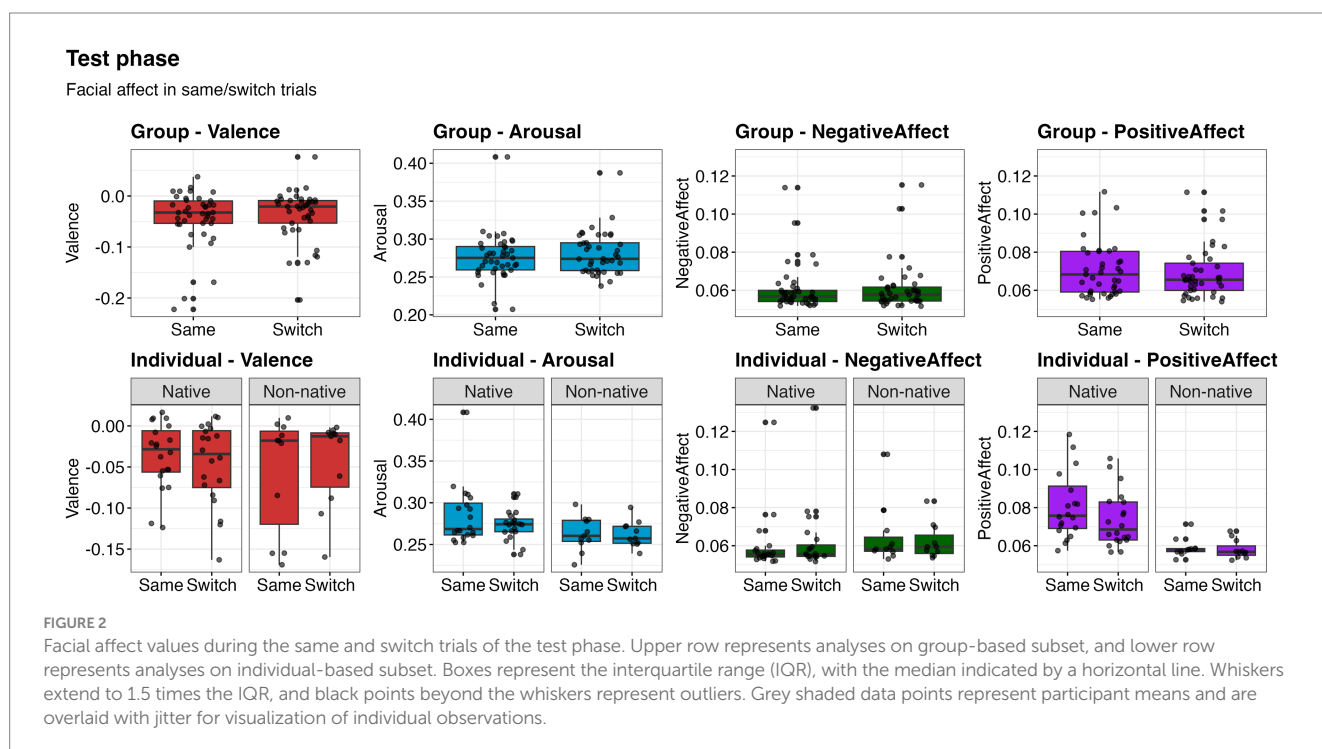
4 Discussion

This study explored infant facial affect during a visual habituation paradigm to examine perceptual attunement to native and non-native speech sounds. By adding facial expression analysis to traditional gaze-based measures, we sought to assess the potential of integrating this measure into the study of perceptual attunement.

Effects of facial affect and arousal were strongest during the habituation phase. Arousal decreased over habituation trials, which is consistent with observations of declining engagement as habituation advances (Hunter and Ames, 1988; Thompson, 2009). Positive affect increased, possibly reflecting a growing preference for familiarized items (Hunter and Ames, 1988). At the same time, global valence as estimated by Baby FaceReader decreased. While we calculated positive affect based on AU12 (lip corner puller), which closely aligns with manual codings of infant positive affect (Zaharieva et al., 2024), the valence score is a composite of several action units that correlates more weakly with manual coding. Future research is needed to establish the relationship of these scores to infant affective expressions.

No significant differences in facial expressions emerged between same and switch trials in the test phase, suggesting that these conditions might not elicit sufficiently strong changes in affective responses.

In our exploratory analysis relating facial expressions to discrimination results, we observed that better discrimination of native contrasts was linked to lower negative affect. This pattern may indicate that successful processing of familiar contrasts leads to reduced frustration or negative responses. Preference for native sounds has also been suggested to reflect expectation of information (Begus et al., 2016), and the present results could indicate that infants who are better in detecting relevant information show better perceptual attunement.



Facial affect and speech sound discrimination

Relationship between affect and novelty scores

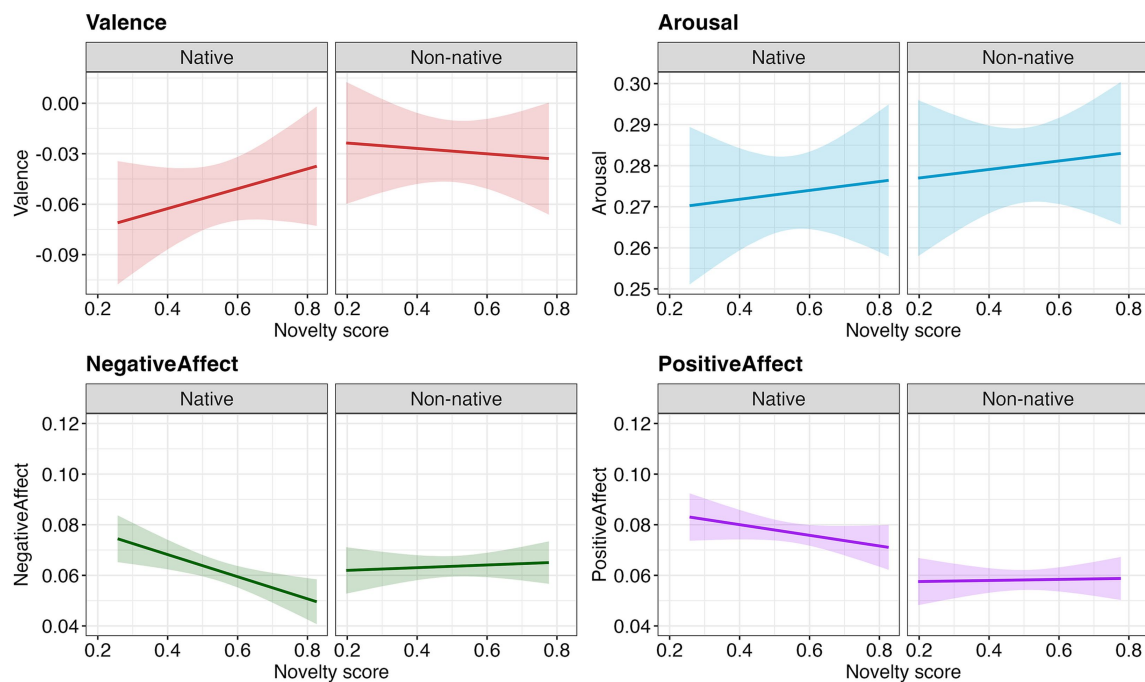


FIGURE 3

Relationship between facial affect in the habituation phase and novelty score (indicator of speech sound discrimination) in the test phase, based on model estimates. The shaded region represents the 95% confidence interval around the fitted values.

Stimulus nativeness affected facial expressions, such that arousal and valence were higher for non-native than native stimuli in the habituation phase, while positive affect was higher for native stimuli in both habituation and test. Higher overall arousal for non-native stimuli might reflect infants' weaker familiarity with those stimuli (Hunter and Ames, 1988). The higher positive affect for native stimuli echos the broader literature showing infants' preferences for native speech sounds (e.g., Jusczyk et al., 1993), but could also be linked to overall higher affective valence of the native stimulus recordings (see stimuli section). The difference in direction of valence and positive affect effects for non-native stimuli reiterates the difficulty of interpreting these. Overall, the differences observed between native and non-native contrasts in the present study are inconclusive, inviting caution in the interpretation of results.

Age-related differences in affective responses were also apparent. For instance, positive affect during habituation was higher for 9-month-old than 12-month-old infants in native, but not non-native trials. These findings might suggest an interplay between developmental changes in perceptual attunement and affective responses.

Overall, this study demonstrates that facial expression analysis can complement traditional measures in studies of perceptual attunement, particularly by capturing affective changes during the habituation phase of laboratory experiments.

This study highlights the potential of facial expression analysis as a tool for tracking the attunement process in laboratory settings. While the learning phase of such studies is typically not the focus of analysis, monitoring facial affect may provide valuable insights into infants' engagement with speech stimuli and the mechanisms that

shape phonetic learning. For instance, in distributional learning studies, infants are exposed to unimodal or bimodal speech sound distributions over a short period, and their ability to categorize speech sounds based on exposure is then assessed (Maye et al., 2002). It has been shown that distributional learning, while still possible outside of sensitive periods, occurs more efficiently within them, requiring less exposure to reach the same learning outcome. Tracking facial expressions during exposure – both inside and outside critical period windows – could provide complementary evidence on the role of affect and arousal in phonetic acquisition during sensitive periods. Similarly, monitoring facial expressions in studies comparing speech sound acquisition from socially enriched versus less enriched stimuli (e.g., Kuhl et al., 2003) could offer new insights into the role of affect and arousal on learning.

Our study also highlighting limitations: not all stimuli elicit detectable differences in facial affect, partly due to the low valence and arousal levels associated with the non-social, non-emotional stimuli in our paradigm. Indeed, the range of valence values in our data largely falls within the neutral/negative range rather than positive (Zaharieva et al., 2024). These constraints likely contribute to the lack of differences in facial expressions during the test phase of our study, must be carefully considered when applying facial affect analysis to studies that are not designed to capture arousal and affect in the first place.

Our findings provide preliminary evidence that facial affect reflects engagement and processing differences during habituation to native and non-native speech sounds. While gaze-based measures remain central to studying perceptual attunement, facial expressions offer a promising, though context-dependent, tool for capturing

affective and cognitive dimensions. The data used in the present study are part of a larger data collection including measurements of infants' socio-communicative environment as well as language outcomes. Relating such measures to facial affect data could deepen our understanding of how infants' affective responses and engagement during speech sound processing are linked to language learning, and may contribute to the ongoing debate on the extent to which speech sound acquisition is constrained by neural plasticity and whether learning mechanisms differ across developmental windows (Werker and Hensch, 2015). This may provide a new avenue for exploring the connections between environmental influences, individual differences, and language development outcomes.

Data availability statement

The anonymized derived data supporting the conclusions of this article, that is, the BabyFaceReader output files and output files from manual coding of looking behavior, will be made available by the authors whenever caregivers have agreed to data sharing in their consent form. Further enquiries should be directed to the corresponding author(s).

Ethics statement

The studies involving humans were approved by Life sciences research ethics and safety, The University of Tokyo. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

ST: Conceptualization, Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. FA: Data curation, Software, Writing – original draft, Writing – review & editing. HH: Data curation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. NK: Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. LP: Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review

References

- Arterberry, M. E., and Bornstein, M. H. (2002). Infant perceptual and conceptual categorization: the roles of static and dynamic stimulus attributes. *Cognition* 86, 1–24. doi: 10.1016/S0010-0277(02)00108-7
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Begus, K., Gliga, T., and Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proc. Natl. Acad. Sci. USA* 113, 12397–12402. doi: 10.1073/pnas.1603261113
- Bosch, L., and Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Lang. Speech* 46, 217–243. doi: 10.1177/00238309030460020801
- Ekman, P., and Friesen, W. V. (1978). Facial action coding system. Palo Alto, CA: Consulting Psychologists Press.
- Fox, J., and Weisberg, S. (2019). An R companion to applied regression. 3rd. Thousand Oaks, CA: Sage. Available online at: <http://z.umn.edu/carbook>
- Hunter, M. A., and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Adv. Infancy Res.* 5, 69–95.
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., and Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *J. Mem. Lang.* 32, 402–420. doi: 10.1006/jmla.1993.1022
- Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9, F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x
- Kuhl, P. K., Tsao, F. M., and Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA* 100, 9096–9101. doi: 10.1073/pnas.1532872100
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

& editing. IL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by JSPS KAKENHI Grant Number 20H05919 and by the MEXT World Premier International Research Center Initiative. BabyFaceReader software was provided by the Centre for Research in Brain, Language and Music.

Acknowledgments

We thank all infants and their parents who have participated in our study, as well as Tomoko Takashima and Rie Ono for their help in recruiting and managing the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. To shorten parts of the article by reformulating passages, and to proofread.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Lenth, R. (2019). Emmeans: estimated marginal means, aka least-squares means. Available online at: <https://cran.r-project.org/web/packages/emmeans/index.html> (Accessed December 20, 2024).
- Lovčević, I., and Tsuji, S. (2024). The developmental pattern of native and non-native speech perception during the 1st year of life in Japanese infants. *Infant Behav. Dev.* 76:101977. doi: 10.1016/j.infbeh.2024.101977
- Maroulis, A. (2018). "Baby FaceReader AU classification for infant facial expression configurations" in *Measuring behavior 2018*. eds. R. A. Grant et al. (UK: Manchester).
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3
- Noldus Information Technology (2021). FaceReader: tool for automatic analysis of facial expressions [computer software]. Wageningen, Netherlands: Noldus Information Technology.
- Oakes, L. M., Sperka, D., DeBolt, M. C., and Cantrell, L. M. (2019). Habit2: a stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behav. Res. Methods* 51, 1943–1952. doi: 10.3758/s13428-019-01244-y
- R Core Team (2013). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Sato, Y., Sogabe, Y., and Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Dev. Psychol.* 46, 106–119. doi: 10.1037/a0016718
- Stager, C. L., and Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388, 381–382. doi: 10.1038/41102
- Thompson, R. F. (2009). Habituation: a history. *Neurobiol. Learn. Mem.* 92, 127–134. doi: 10.1016/j.nlm.2008.07.011
- Tsuji, S., and Cristia, A. (2014). Perceptual attunement in vowels: a meta-analysis. *Dev. Psychobiol.* 56, 179–191. doi: 10.1002/dev.21179
- Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., et al. (1994). Discrimination of English /r-/ and /w-y/ by Japanese infants at 6–12 months: language-specific developmental changes in speech perception abilities. *Proceedings 3rd International Conference Spoken Language Process (ICSLP 1994)* 1695–1698.
- Werker, J. F., and Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annu. Rev. Psychol.* 66, 173–196. doi: 10.1146/annurev-psych-010814-015104
- Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3
- Wickham, H., Chang, W., and Wickham, M.H. (2016). Package 'ggplot2': Create elegant data visualisations using the grammar of graphics. Available online at: <https://cran.r-project.org/web/packages/ggplot2/index.html> (Accessed December 20, 2024).
- Zaharieva, M. S., Salvadori, E. A., Messinger, D. S., Visser, I., and Colonnesi, C. (2024). Automated facial expression measurement in a longitudinal sample of 4- and 8-month-olds: baby FaceReader 9 and manual coding of affective expressions. *Behav. Res. Methods* 56, 5709–5731. doi: 10.3758/s13428-023-02301-3